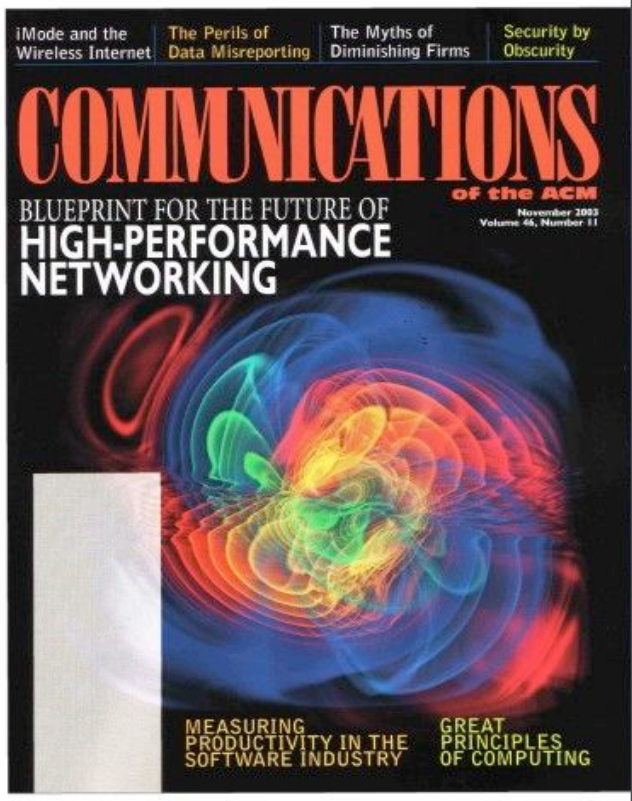
CyberInfrasructure Update

and miscellaneous remarks

Edward Seidel

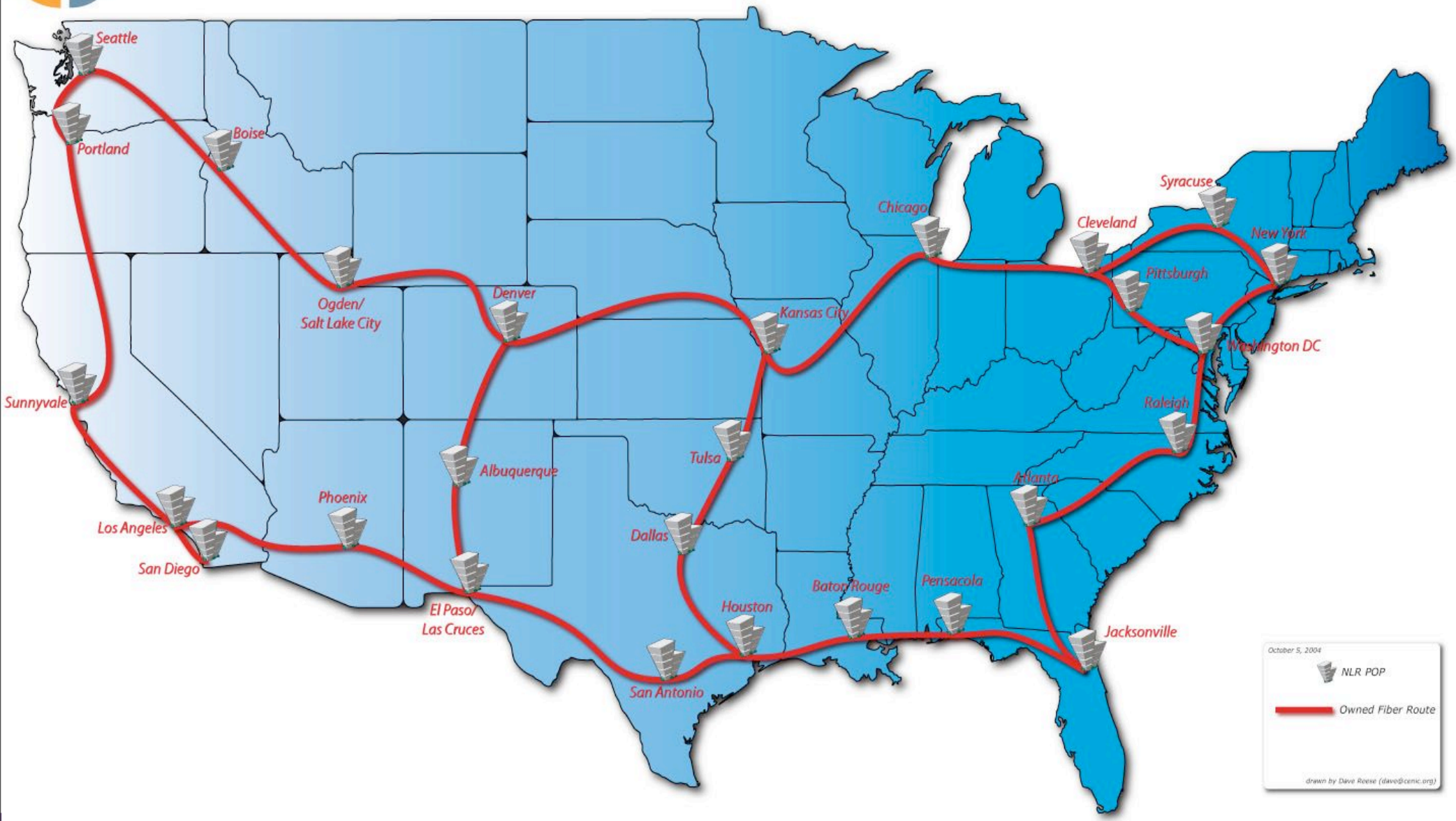
CCT Director

LSU

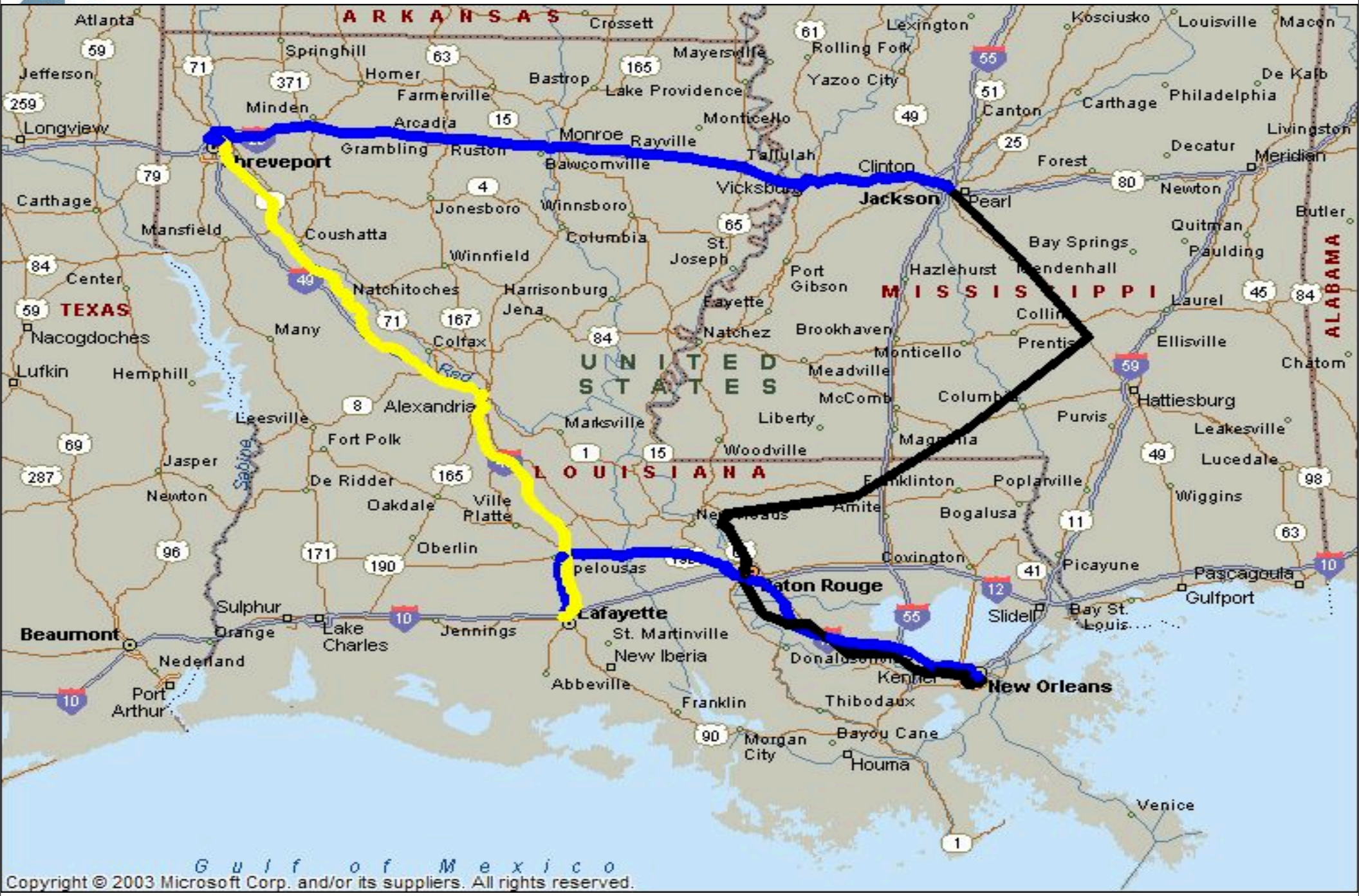


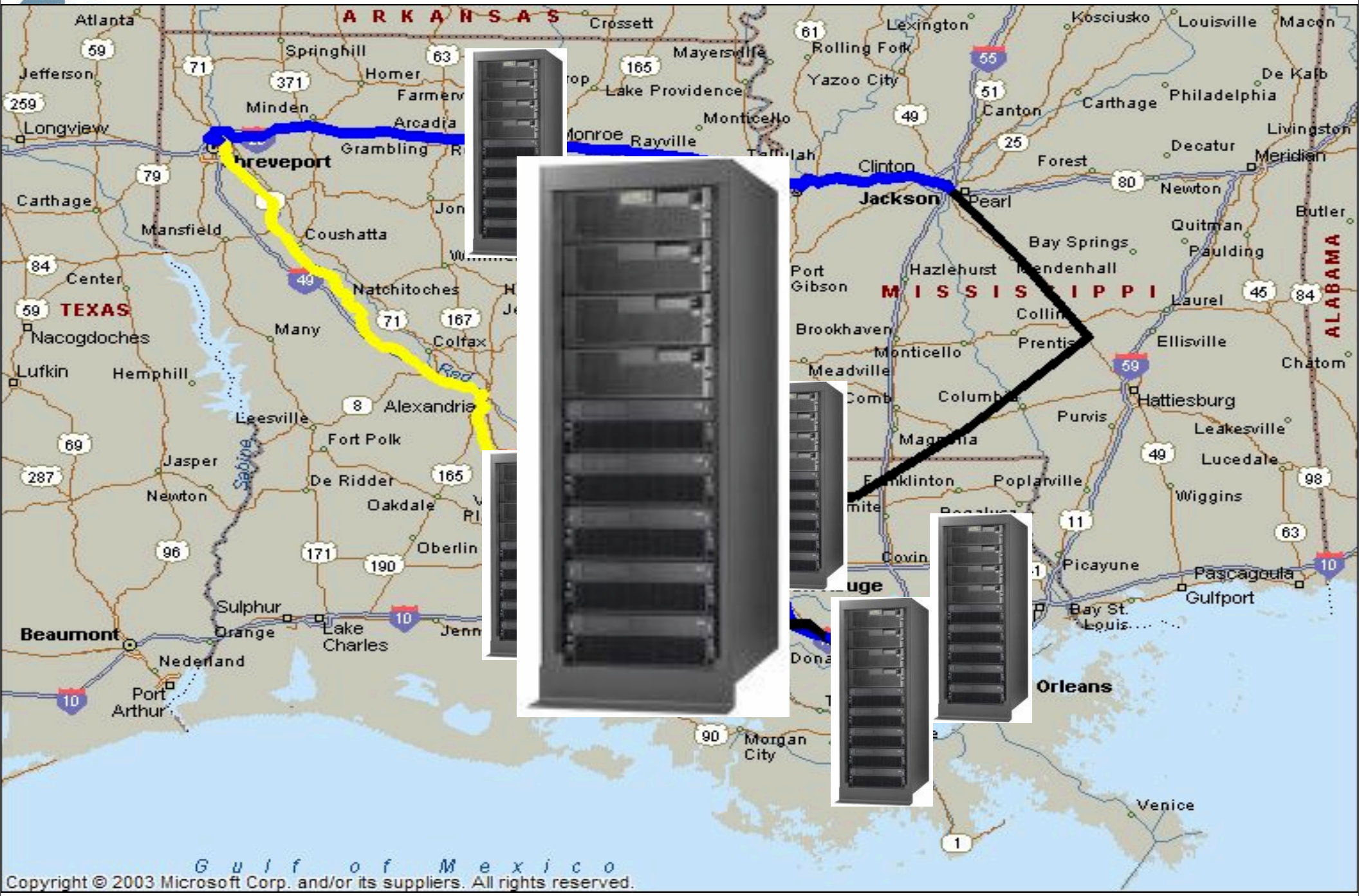
NLR
light the future

National LambdaRail Architecture



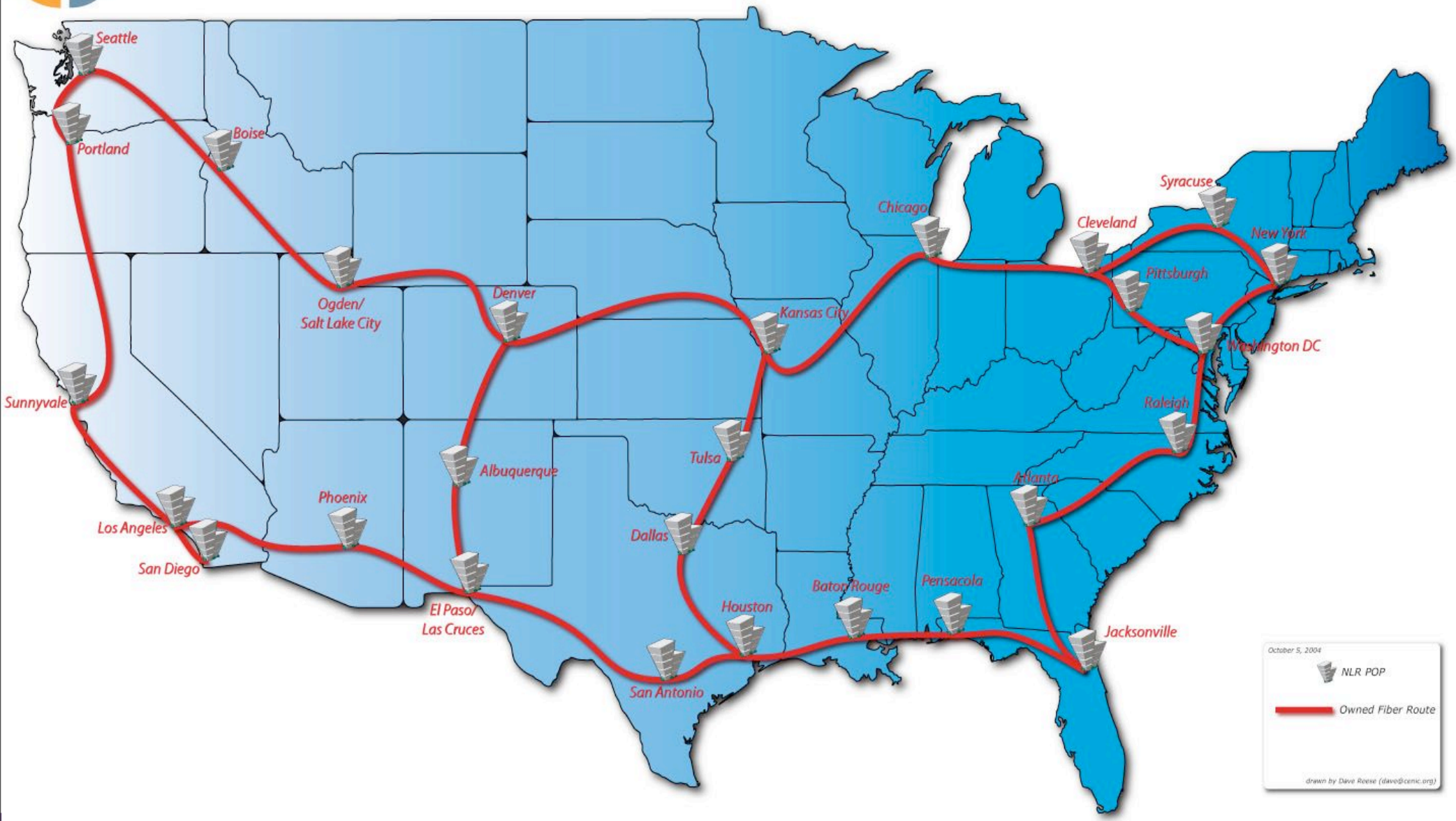
For more information regarding NLR see http://www.nlr.net or contact info@nlr.net





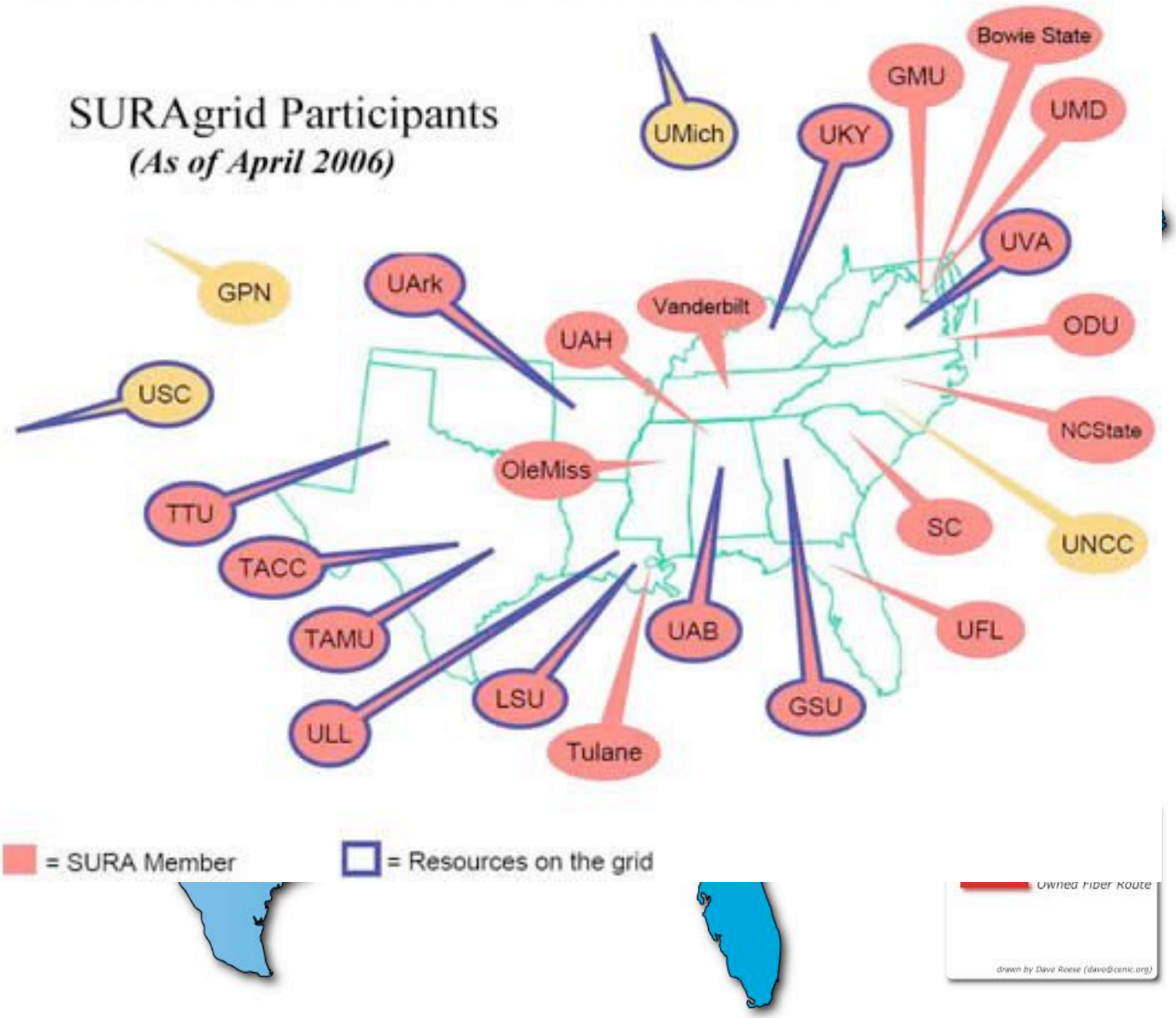
NLR
light the future

National LambdaRail Architecture



For more information regarding NLR see http://www.nlr.net or contact info@nlr.net

National LambdaRail Architecture



For more information regarding NLR see http://www.nlr.net or contact info@nlr.net

The Power of RONs

- LA's Vision 20/20: $25M annually across 5 campuses
- Louisiana Optical Network Initiative (LONI)
  - $50M: Gov. Blanco investment ripples across region
  - NLR pathway changed, BoR membership
  - Creating most advanced state infrastructure in the US
  - New 100 TF commitment: biggest facility in nation, "equal" to current NSF facilities combined!
  - LSU, Tulane, University of New Orleans, University of Louisiana-Lafayette, Louisiana Tech, Southern University, Community College system
  - Louisiana Economic Development Authority
- LONI Institute
  - $30M BoR proposal, 6 Universities, CS, Materials, Bio

Creating a Regional Environment

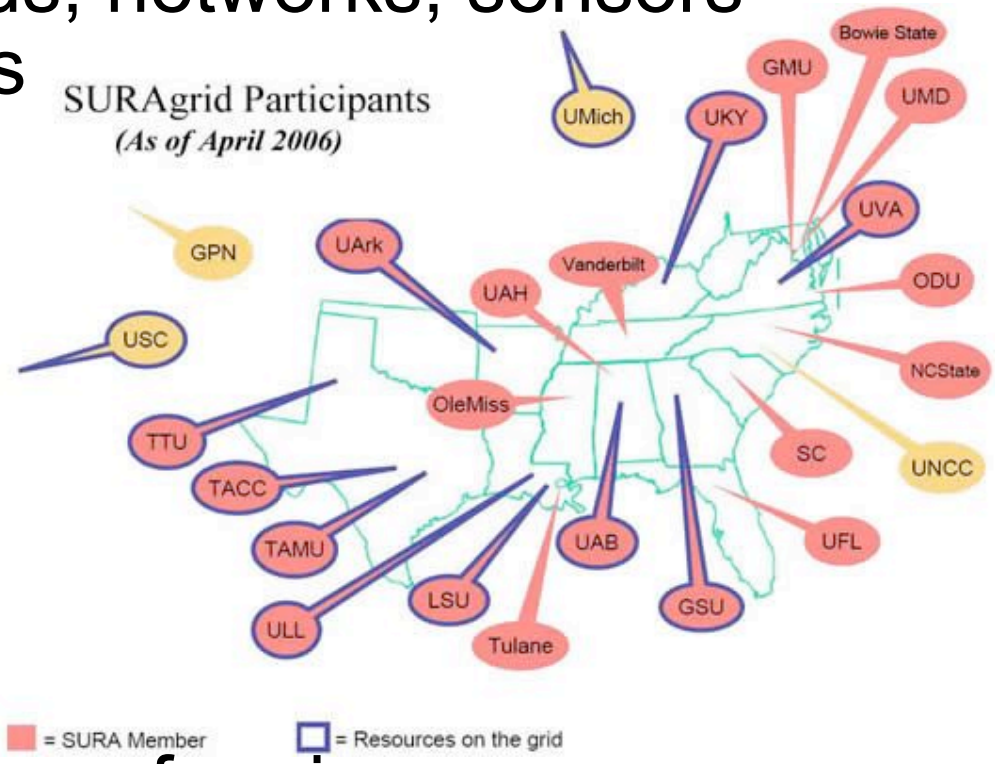
- Suddenly collaborative, competitive
  - meetings, collaborations between universities
- Many projects now proposed on top of LONI
  - NSF ESPCOR: CyberTools
    - Common software infrastructure for MD, CFD, Experiment
    - Viz services, data services, computing services, co-scheduling, portals
    - Outreach down to high schools
  - DOE EPSCPR: UCoMS
  - NSF MRI: PetaShare, VizTangibles, VONET
  - NIH LBRN, UCoMS, Enlightened, HD Classes
  - NSF Center proposals
    - Track 1, Track 2, HPCOPS

Perhaps a good model for SURAgrid

- Leverage the collaboration, infrastructrure for community enabling projects
- Biggest problem: precious few know what to do
  - Board of Regents proposal
  - SURAgrid can have a lot of impact

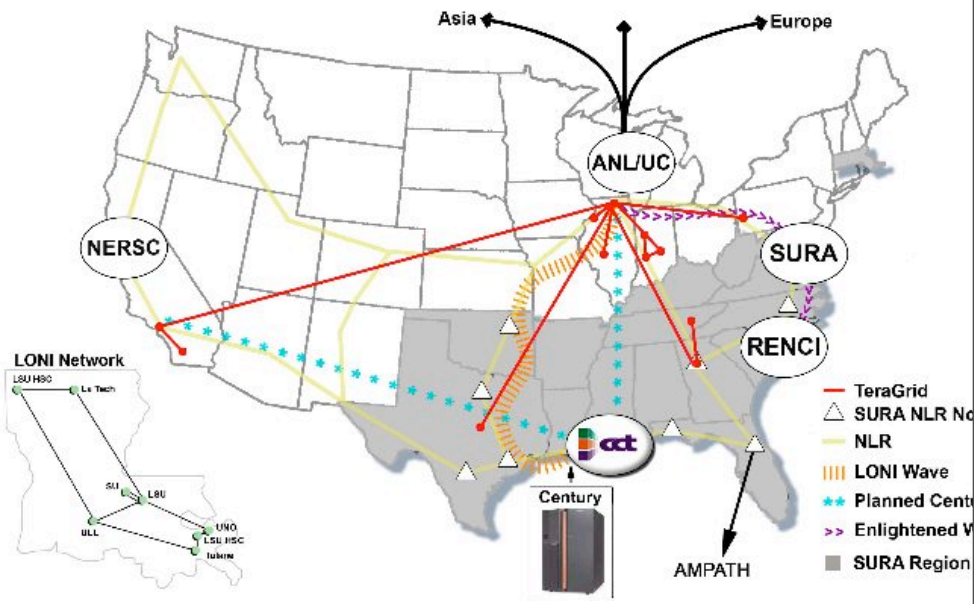
SURAgrid as part of National CI

- Specific Apps in Region: from coastal ecology to diabetes, 21st century research, driven by advances in computing
- High-end computing, grids, networks, sensors critical to these advances
- SURA has excellent network infrastructure
- BUT, SURA regional usage is just 14% of NSF facilities
- We can leverage strengths to advance science of region



Recent Developments

- SURA assets in networking, NLR nodes
- SURAgrid, GridPlan
  - Organizing activities, many projects, supporting SCOOP, funded proposals coming
  - Still lacking strong HPC component, but coming with IBM, DEll, Sun, other HPC initiatives
- NSF Petascale Initiatives
  - Track 1, Track 2
  - Leverage SURA investments, networks, grids, for outreach

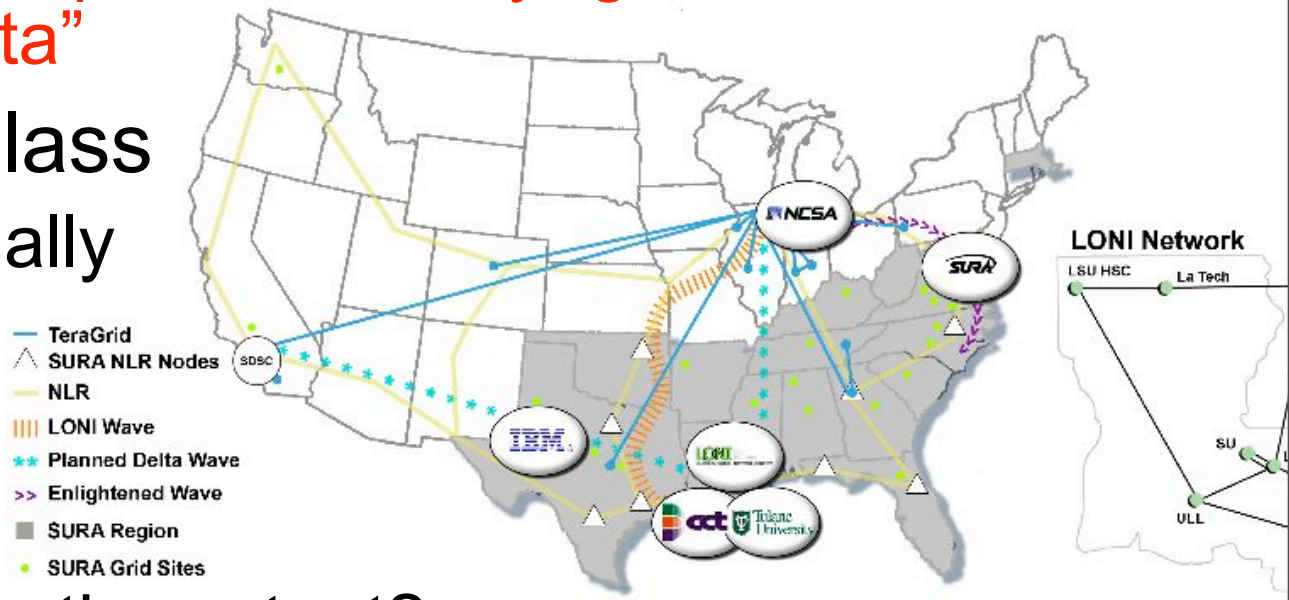


Basic Ideas

- Strengthen SURAgrid to create *the* leading regional HPC environment + *program to develop/use it*
  - Deploy numerous *integrated* regional supercomputers
  - Leverage regional investments in optical networks
    - SURA, NLR, RONs: Gbit makes regional/national integration work
  - Coordinate deployment, operations, file systems, scheduling, cycle sharing, training, align with Tgrid, national facilities, event-driven, etc
- Major impact across region
  - Raise level of science, engineering, and humanities; support specific projects: SCOOP, DynaCode, Event Drive applications
  - Increase use of HPC, IT: Nation:SURA = 3.5:1
  - Alignment with NSF cyberinfrastructure plan, position for funding opportunities: Read this!

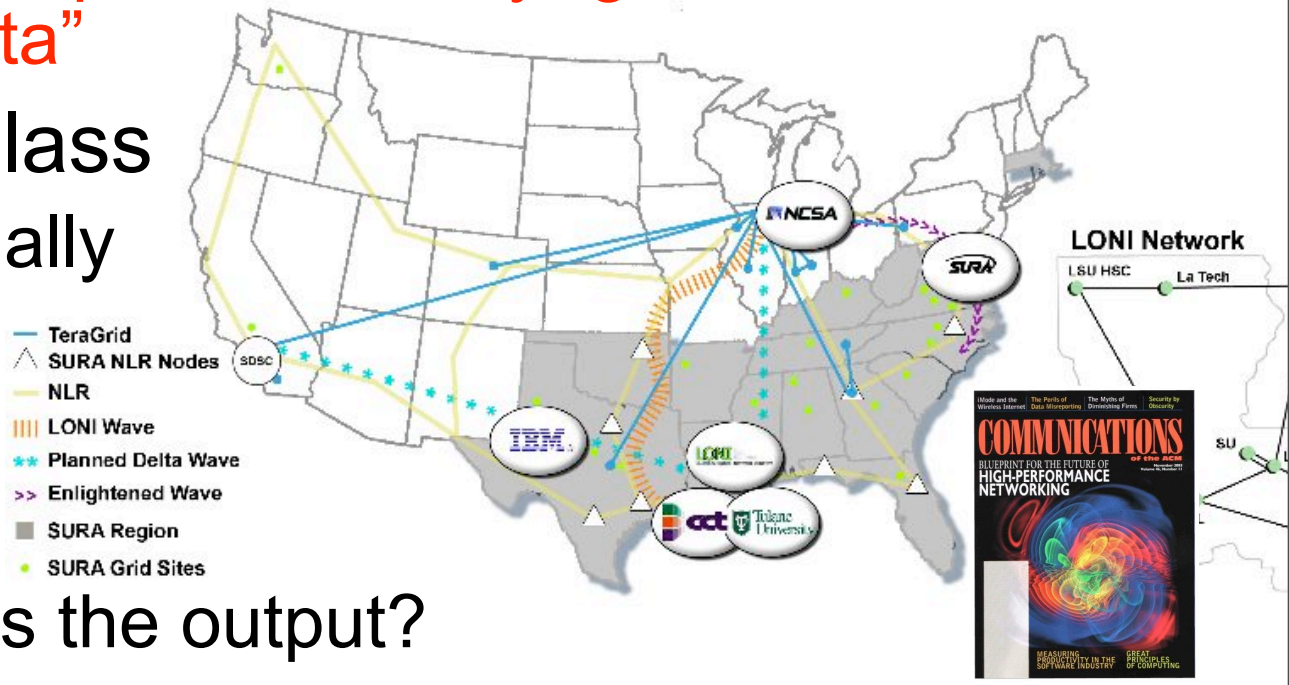
Really Big Supercomputers are Coming

- NSF "Toward Petascale" (Track 2)
  - LSU "Delta": November, 2006, 3x earlier "Century" proposal in just a year
    - LSU/LONI-led, NCSA, SURA, Tulane, LA Tech
    - Typically have 500TF, 20-40K procs, 100TB
    - Data! "Supercomputers "merely" generators of petabytes of data"
- NSF Leadership Class
  - 4 proposals nationally
  - 500K-1M procs
  - 10-20PF, 1-5 PB
  - How we will access the output?



Really Big Supercomputers are Coming

- NSF "Toward Petascale" (Track 2)
  - LSU "Delta": November, 2006, 3x earlier "Century" proposal in just a year
    - LSU/LONI-led, NCSA, SURA, Tulane, LA Tech
    - Typically have 500TF, 20-40K procs, 100TB
    - Data! "Supercomputers "merely" generators of petabytes of data"
- NSF Leadership Class
  - 4 proposals nationally
  - 500K-1M procs
  - 10-20PF, 1-5 PB
  - How we will access the output?



Petascale Challenges for GR *(Scary)*

- Basic numerical algorithms in place (new!)
  - Dozens of groups compute GR, share common tools for petascale
- Most AMR codes scale to ~ 64 processors, but petascale machine will have 500K processors!
  - Serious work to do on scaling, dynamic loadbalancing, fault tolerance, software environments, debugging tools
- Consider novel ways to break up jobs
  - Farming off non-parallelizable analysis tasks, viz, coupling codes, spawning, integrating with data analysis
  - Consider petascale machine as your grid!
- Handling all the data: storing, retrieving, visualizing, analyzing. Lambda provisioning for steering, migration, viz
- Parameter space! Metadata to describe 10^{6+} simulations
- SURA can play an important role

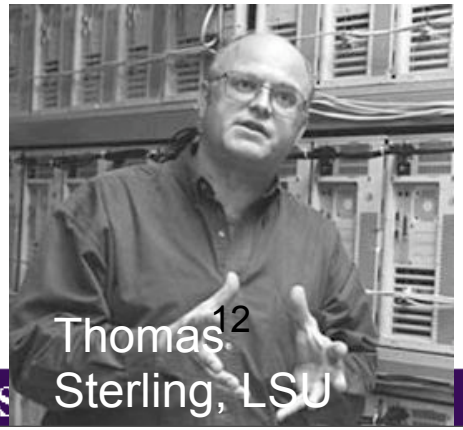
SURA Role in Track 1, Track 2

- SURAgrid is a well organized, proactive organization: 60+ universities!
- Very dense optical network infrastructure
- How to leverage?
  - Work with centers (e.g., Track 1, Track 2...)
  - Develop SURAgrid HPC resources, align with centers: hardware, CI, libraries, logins, etc
  - Use local machines as “outposts”
    - training, analysis, development
    - easy to scale up to national centers
    - increase userbase, address 3.5:1 ratio
  - Used this idea in NSF proposals

Optical Nets Enable Old & New Paradigms

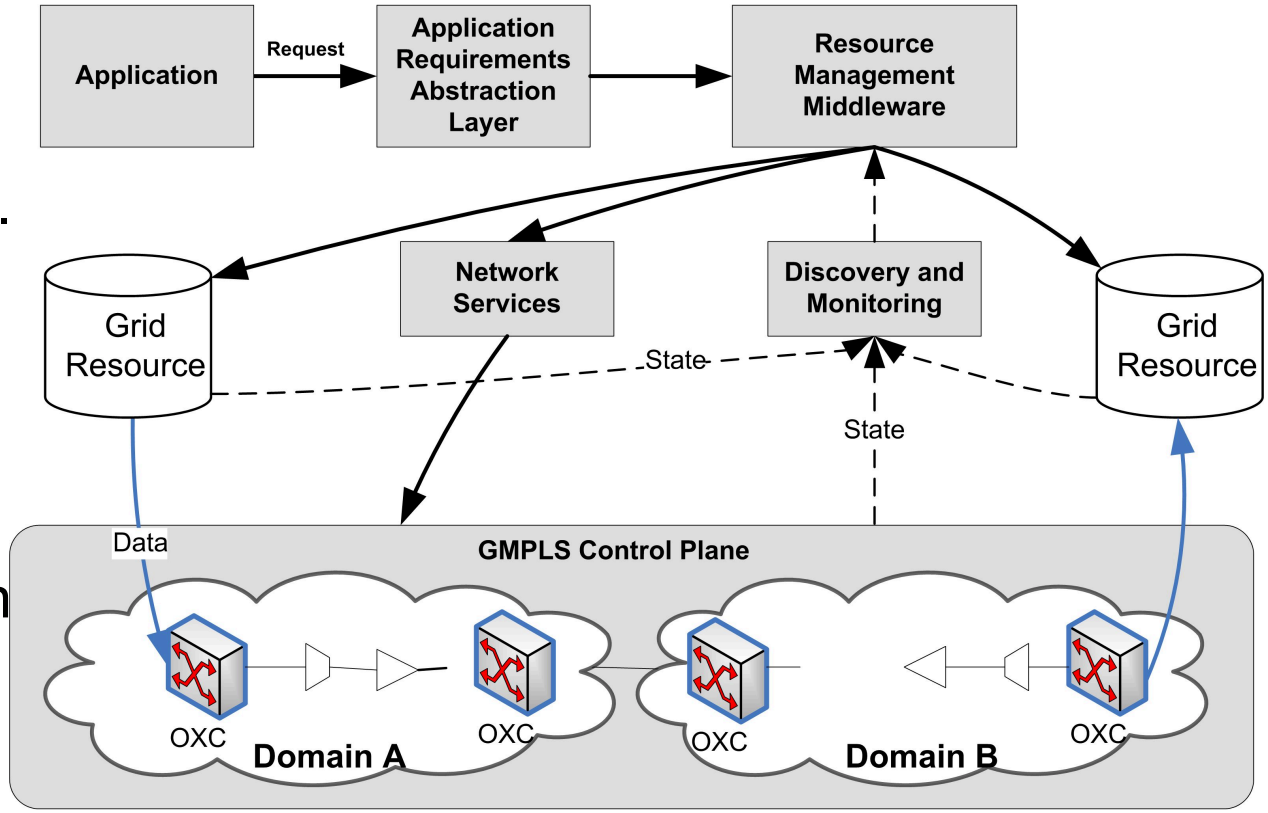
Huge BH simulation, outputs ~TBs data, must be interactively viz'd & analyzed by international collaboration

- Many practical problems found above go away!
- New Distributed Computing, Data, and Viz Services
  - Data larger than memory, disk slow: multiple streaming data servers
  - Co-scheduling machines, data, and network for single event: MacLaren HARC
  - Lambda Provisioning: Karmous-Edwards EnLIGHTened NSF project
  - Same application must work with multiple nets, services: GAT/SAGA
- High Resolution: not science w/o details, analysis
  - Uncompressed HD stream: near latency-free remote viz at 1.5Gbit/sec
  - Video for colleagues: see details, read equations
  - Sterling HPC class: from LSU but LA, AR, Brno (Czech)



enλIGHTened computing

- Develop ability for network reconfiguration, on-demand, in-advance reservation. Intelligent provisioning of lambdas.
- Develop a Grid framework for applications/users to request/coordinate *any* resources
- Abstractions for application developers to do this!
- Ensure resources are successfully scheduled: *Highly-Available Robust Co-Scheduler* (HARC)



This is already being used in Sterling's class!

Spinoff: HD Video for Distance Learning – An Introduction to HPC

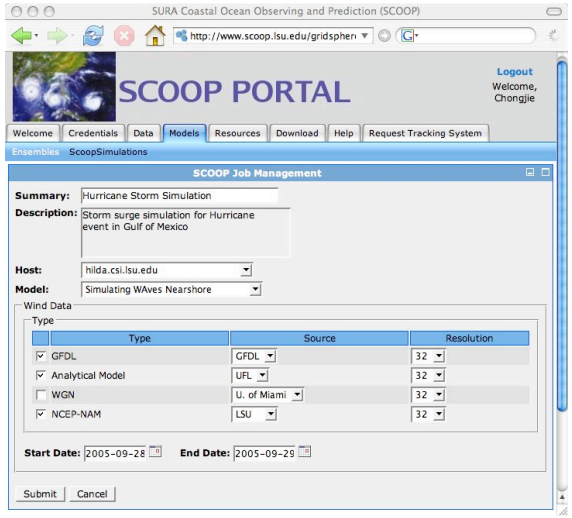
- Thomas Sterling: Introduction to all aspects of Supercomputing
- In collaboration
  - U. Arkansas, Louisiana Tech, Masaryk (Czech), MCNC
  - Taken for local credit at other sites
- Multimedia
  - Web site for all course materials and video lectures
  - Easily accessible for review/study
  - Entire class fits on iPod
- High Definition video over Internet
  - Uncompressed HD (why? latency!)
  - 1920x1080 resolution, 60 frames per second interlaced (1080i standard)
  - Capturing HD-SDI video w/Centaurus cards
  - RTP (Real-time Transport Protocol) for video payload, uncompressed video data
  - 1.5 Gbit/sec: need to experiment w/ compression

Spinoff: HD Video for Distance

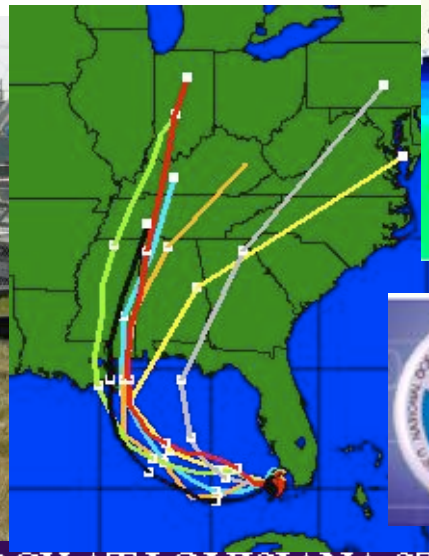
compression

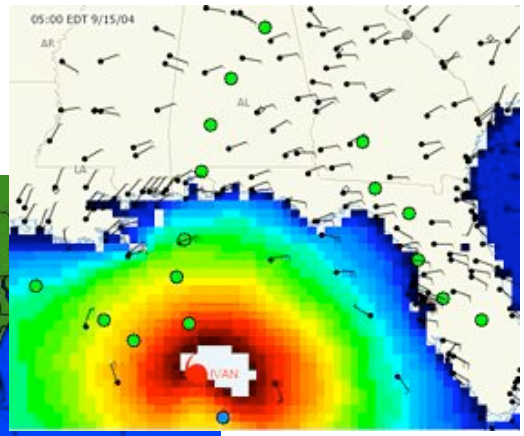
SURA Coastal Ocean Observing Program (SCOOP)

- Integrating data from regional observing systems for realtime coastal forecasts in SE
- Coastal modelers working closely with computer scientists to couple models, provide data solutions, deploy ensembles of models on the Grid, assemble realtime results with GIS technologies.



University of Alabama at Huntsville, University of Florida, GoMOOS, Louisiana State University, University of Miami, University of Maryland, University of North Carolina, Texas A&M, Virginia Inst of Marine Sciences

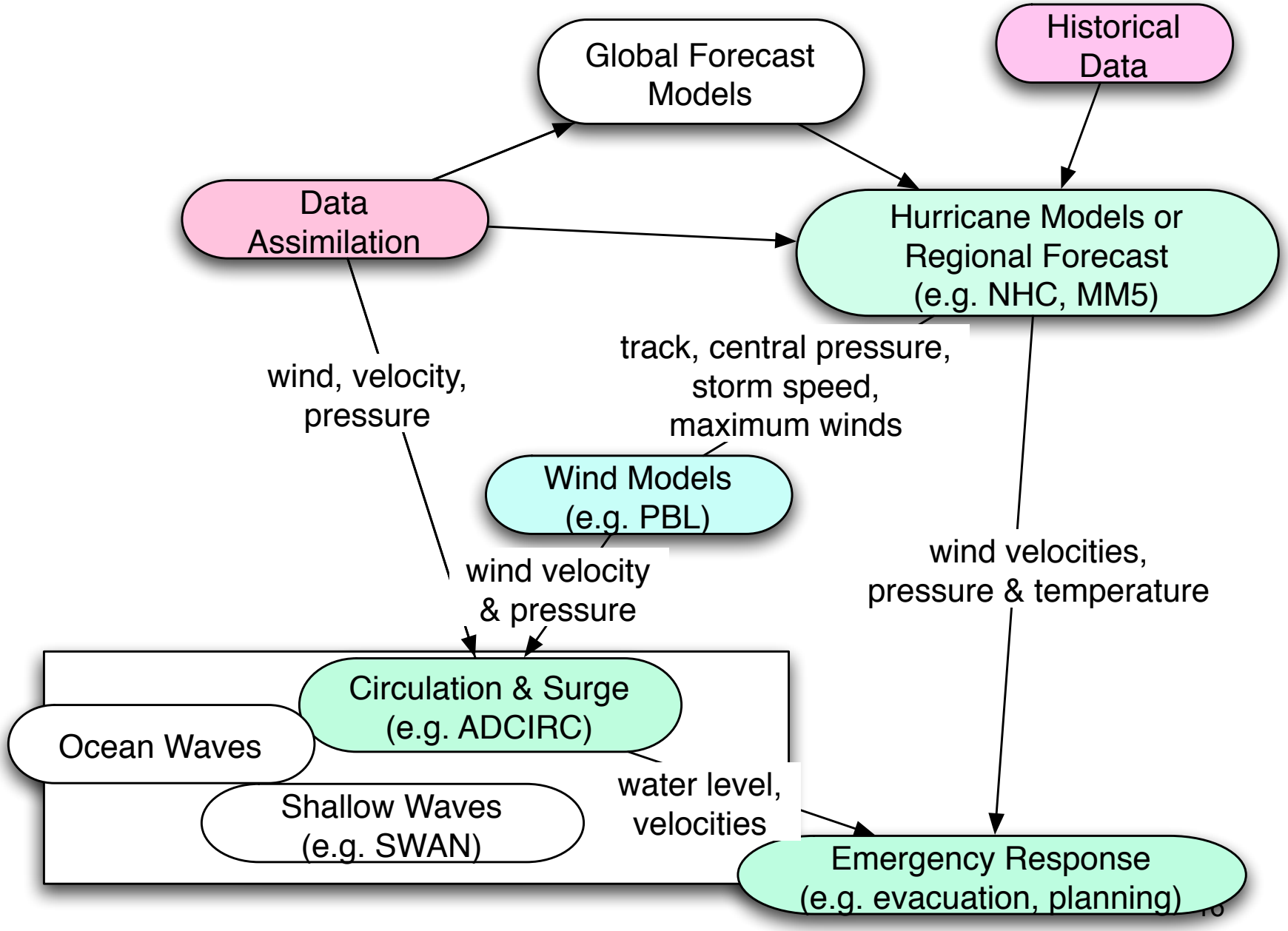




Office of Naval Research

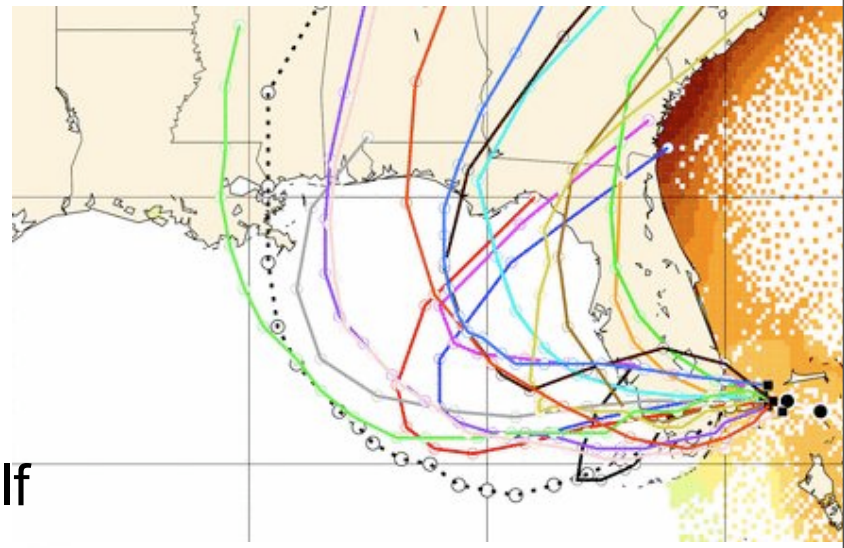
CENTER FOR COMPUTATION & TECHNOLOGY AT LOUISIANA STATE UNIVERSITY

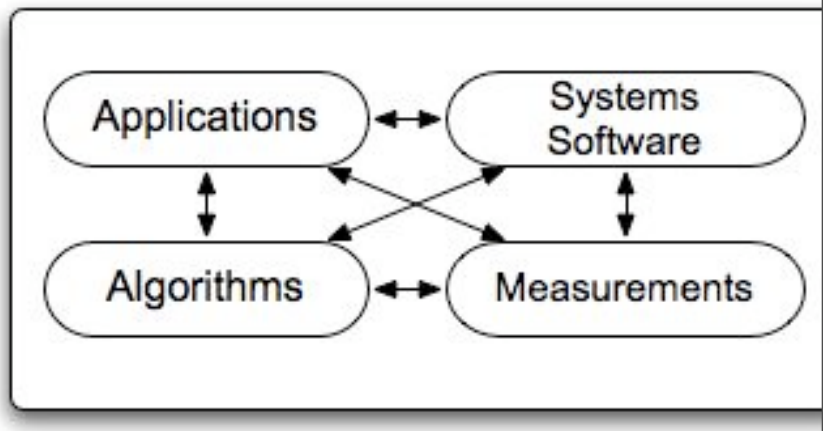
Model-Model-Data Coupling



DynaCode

- Focus on scenarios:
  - Hurricane ensemble modeling
    - Coupling ocean circulation, storm surge, wave generation models for Gulf
    - Notifications from NHC trigger customized ensemble hurricane model (surge/wind/wave), sensors verify, guid dynamic ensembles
    - Event driven, dynamic component framework with algorithm selection, optimization tools, workflow, data assimilation, result validation with sensor/satellite.
  - Ecological restoration and control
    - Coupled models (hydrodynamic, salinity, sediment) control diversion, wind fields inject real time data.





Event-driven Computing

- Regional applications: hurricane, storm surge very important
- Could develop event-driven computing as a specialty
- Serious interest nationally in many communities: hurricanes, earthquakes, disease outbreak, chemical plant explosions (disasters of all kinds...)
- Could be an excellent focus for SURAgrid research: pioneering regional project

Summary

- National, regional optical networks, petascale machines provide opportunities
- Agencies investing in CI
- SURA can play a major role
  - harness its optical net strengths, SURAgrid
  - outposts of national centers
  - regional application development
  - SCOOP as example of event-driven
  - educational applications of optical nets: SURA-wide class?